Learning reduced models for motion estimation on long temporal image sequences

Isabelle Herlin, Karim Drifi

LEARNING REDUCED MODELS FOR MOTION ESTIMATION ON LONG TEMPORAL IMAGE SEQUENCES

Isabelle Herlin[1,2]*, Karim Drifi*[1,2]

[1]INRIA, Institut National de Recherche en Informatique et Automatique,
Domaine de Voluceau, Rocquencourt - BP 105, 78153 Le Chesnay Cédex, France

[2]CEREA, joint laboratory ENPC - EDF R&D,
Université Paris-Est, 6-8 avenue Blaise Pascal,
Cité Descartes, Champs-sur-Marne, 77455 Marne la Vallée Cédex 2, France

ABSTRACT

***Index Terms*—** Motion, Model Reduction, Principal Order Decomposition, Model Coupling, Galerkin Projection

1. INTRODUCTION

Motion estimation from an image sequence has been intensively studied since the beginning of image processing [1, 2]. The aim is to retrieve the velocity field $\mathbf{w}(\mathbf{x},t)$ visualized by a discrete image sequence $I = \{I^z\}_{z=1...Z} = \{I(\mathbf{x},t_z)\}_{z=1...Z}$. The application of data assimilation techniques to motion estimation also emerged in the last five years [3, 4, 5]. Data assimilation aims to find the optimal solution to the equations describing the temporal evolution of motion and image fields and to that, named observation equation, which links the motion field to the observed image acquisitions. Its advantage, compared to image processing methods, is the retrieval of a continuous motion field, allowing to assess the processes occurring between two acquisitions. Its major drawbacks are the memory and computer resources required when using the full model, defined on the pixel grid. This does not allow to process long temporal sequences of large size images such as long-term satellite data. To get round this problem, reduction methods are required to apply data assimilation on subspaces. In [6], such reduced model is proposed to describe the dynamics of motion and image fields. Coefficients characterizing observations in the image subspace are then assimilated in the reduced model to estimate those characterizing the motion field.

In this paper, we focus on the estimation of motion on a long temporal window. A method, named sliding windows assimilation, is designed: image assimilation is first applied on an initial sub-window with the pixel grid, and then on sliding temporal windows with reduced models, obtained by Galerkin projection on motion and image subspaces. The full model, used on the pixel grid in this paper, is described in Section 2. The main concepts of data assimilation are briefly summarized in the same section. The derivation of the reduced model by Galerkin projection is then explained in Section 3. Last, the sliding windows assimilation method is fully described in Section 4 and its results quantified in Section 5.

2. FULL MODEL DESCRIPTION AND DATA ASSIMILATION

In order to describe the sliding windows assimilation method, we consider, as an example, divergence-free motion fields $\mathbf{w}(\mathbf{x},t)$, with $\mathbf{w} = \begin{pmatrix} u & v \end{pmatrix}^T$, $\mathbf{x} = \begin{pmatrix} x & y \end{pmatrix}^T \in \Omega$, a bounded domain, and $t \in [t_0, t_N]$, a closed interval. However, the discussion would be still valid for other motion dynamics. As a divergence-free motion field $\mathbf{w}(\mathbf{x},t)$ is fully characterized by its vorticity value $\xi(\mathbf{x},t)$, $\mathbf{w}$ is a function of ξ, and written $\mathbf{w}(\xi)$. We also assume that the motion field satisfies the heuristics of Lagrangian constancy described by: $\dfrac{d\mathbf{w}}{dt} = \dfrac{\partial \mathbf{w}}{\partial t} + (\mathbf{w}^T \nabla)\mathbf{w} = 0$. Under the divergence-free assumption, this is rewritten as an equation describing the evolution of ξ:

$$\frac{\partial \xi}{\partial t} + \mathbf{w}(\xi) \cdot \nabla \xi = 0. \tag{1}$$

The state vector of the full model is defined as the function: $\mathbf{X}(\mathbf{x},t) = \begin{pmatrix} \xi(\mathbf{x},t) & I_s(\mathbf{x},t) \end{pmatrix}^T$. I_s is a pseudo-image, which has the same dynamics as the image observation: the motion field transports it according to the so-called optical flow equation [1]:

$$\frac{\partial I_s}{\partial t} + \mathbf{w}(\xi) \cdot \nabla I_s = 0. \tag{2}$$

This pseudo-image I_s is included in the state vector in order to allow an easy comparison with the image observations at each acquisition date: they have to be almost identical.

Eqs. 1 and 2 are summarized as:

$$\frac{\partial \mathbf{X}}{\partial t} + \mathbb{M}(\mathbf{X}) = 0 \tag{3}$$

Data assimilation aims to find an optimal solution to the evolution equation (Eq. 3) and to the observation equation that links the state vector to the image observations $I(\mathbf{x}, t)$:

$$\mathbb{H}\mathbf{X} = I \tag{4}$$

The observation operator $\mathbb{H}$ projects the state vector into the space of observations. In our case, $\mathbb{H}$ reduces to a projection on the pseudo-image component, $\mathbb{H}\mathbf{X} = I_s$, and Eq. 4 rewrites as: $I_s = I$.

Image observations are assimilated in the full model $\mathbb{M}$, defined by Eqs. 1 and 2, to estimate vorticity and motion. Full description of the data assimilation method is given in [5].

3. GALERKIN PROJECTION

The aim is to define a reduced model of the full model described in Section 2. We assume that we have knowledge of a rough estimation ξ_b, named background value, of vorticity at the beginning of the temporal window. The first issue is to define subspaces for vorticity and image fields, onto which the evolution equations 1 and 2 are respectively projected to obtain the reduced model. These subspaces are defined by their orthogonal basis Ψ_ξ and Ψ_I, obtained as explained below. First, a Proper Orthogonal Decomposition transform (POD) [7] is applied to the image observations $I = \{I^z\}_{z=1\dots Z}$, that defines Ψ_I^f. Second, ξ_b is integrated in time with Eq. 1. This simulation provides snapshots, on which POD is applied to obtain Ψ_ξ^f. We keep the first K modes of Ψ_ξ^f and the first L modes of Ψ_I^f to define the reduced basis Ψ_ξ and Ψ_I. K and L are chosen in order to keep 95% of variance.

Let $a_i(t)$ and $b_j(t)$ be the projection coefficients of $\xi(\mathbf{x}, t)$ and $I_s(\mathbf{x}, t)$ on Ψ_ξ and Ψ_I. $\xi(\mathbf{x}, t)$ and $I_s(\mathbf{x}, t)$ are approximated by: $\xi(\mathbf{x}, t) \approx \sum_{i=1}^{K} a_i(t)\psi_{\xi,i}(\mathbf{x})$ and $I_s(\mathbf{x}, t) \approx \sum_{j=1}^{L} b_j(t)\psi_{I,j}(\mathbf{x})$. After replacing in Eqs. 1 and 2, it comes:

$$\begin{cases} \displaystyle\sum_{i=1}^{K} \frac{da_i}{dt}(t)\psi_{\xi,i}(\mathbf{x}) + \\ \mathbf{w}\left(\displaystyle\sum_{i=1}^{K} a_i(t)\psi_{\xi,i}(\mathbf{x})\right) \cdot \nabla \left(\displaystyle\sum_{i=1}^{K} a_i(t)\psi_{\xi,i}(\mathbf{x})\right) = 0 \\ \displaystyle\sum_{j=1}^{L} \frac{db_j}{dt}(t)\psi_{I,j}(\mathbf{x}) + \\ \mathbf{w}\left(\displaystyle\sum_{i=1}^{K} a_i(t)\psi_{\xi,i}(\mathbf{x})\right) \cdot \nabla \left(\displaystyle\sum_{j=1}^{L} b_j(t)\psi_{I,j}(\mathbf{x})\right) = 0 \end{cases} \tag{5}$$

This system is projected on each vector of Ψ_ξ and Ψ_I to derive:

$$\begin{cases} \dfrac{da_k}{dt}(t)\left\langle \psi_{\xi,k}, \psi_{\xi,k} \right\rangle + \\ \left\langle \mathbf{w}\left(\sum_{i=1}^{K} a_i(t)\psi_{\xi,i}\right) \cdot \nabla \left(\sum_{i=1}^{K} a_i(t)\psi_{\xi,j}\right), \psi_{\xi,k} \right\rangle = 0, \\ \dfrac{db_l}{dt}(t)\left\langle \psi_{I,l}, \psi_{I,l} \right\rangle + \\ \left\langle \mathbf{w}\left(\sum_{i=1}^{K} a_i(t)\psi_{\xi,i}\right) \cdot \nabla \left(\sum_{j=1}^{L} b_j(t)\psi_{I,j}\right), \psi_{I,l} \right\rangle = 0, \end{cases} \tag{6}$$

$\langle ., . \rangle$ being the scalar product: $\langle f, g \rangle = \int f(\mathbf{x})g(\mathbf{x})d\mathbf{x}$.

As $\mathbf{w}$ is a linear function of the vorticity ξ, System (6) is simplified as:

$$\begin{cases} \dfrac{da_k}{dt}(t) + a^T(t)B(k)a(t) = 0, \quad k = 1 \dots K. \\ \dfrac{db_l}{dt}(t) + a^T(t)G(l)b(t) = 0, \quad l = 1 \dots L. \end{cases} \tag{7}$$

with:

- $a(t) = \begin{pmatrix} a_1(t) & \dots & a_K(t) \end{pmatrix}^T$ and $b(t) = \begin{pmatrix} b_1(t) & \dots & b_L(t) \end{pmatrix}^T$,
- $B(k)$ a $K \times K$ matrix: $B(k)_{i,j} = \dfrac{\langle \mathbf{w}(\psi_{\xi,i}) \cdot \nabla \psi_{\xi,j}, \psi_{\xi,k} \rangle}{\langle \psi_{\xi,k}, \psi_{\xi,k} \rangle}$,
- $G(l)$ a $K \times L$ matrix: $G(l)_{i,j} = \dfrac{\langle \mathbf{w}(\psi_{\xi,i}) \cdot \nabla \psi_{I,j}, \psi_{I,l} \rangle}{\langle \psi_{I,l}, \psi_{I,l} \rangle}$

Let $\mathbf{X}_R(\mathbf{x}, t) = \begin{pmatrix} a(t)^T & b(t)^T \end{pmatrix}^T$ be the state vector of the reduced model. System (7) is rewritten as:

$$\frac{d\mathbf{X}_R}{dt} + \mathbb{M}_R(\mathbf{X}_R) = 0, \tag{8}$$

$\mathbb{M}_R$ being the projection [8] of the full model on Ψ_ξ and Ψ_I.

The same data assimilation method than that described in Section 2 is then applied. The state vector is $\mathbf{X}_R$, the evolution equation Eq. 3 is replaced by Eq. 8, and the components $b_j(t)$ of the state vector are compared with the projections of image acquisitions on $\psi_{I,j}$ in order to implement Eq. 4.

4. SLIDING WINDOWS ASSIMILATION METHOD

This section describes the sliding windows assimilation method, which is defined to process long temporal image sequences.

The discrete sequence $I = \{I^z\}_{z=1\dots Z}$ is first split into short sub-sequences, that half overlap in time. The corresponding temporal intervals or windows are denoted Wi_m, with m being their index. Images belonging to Wi_1 are assimilated in the full model $\mathbb{M}$ described in Section 2. This allows the retrieval of the vorticity on the whole interval Wi_1. Its value at the beginning of Wi_2 is taken as the background value ξ_b, required to define the reduced model $\mathbb{M}_R^2$ on Wi_2,

as explained in Section 3. The coefficients of projection $b_j(t)$, of images belonging to Wi_2, are assimilated in this reduced model in order to retrieve the vorticity coefficients $a_i(t)$, and compute the vorticity values $\xi(t)$ over Wi_2. This again provides the background value of vorticity for Wi_3 and allows to define the reduced model $\mathbb{M}_R^3$ on the third window. The process is then iterated, until the whole image sequence I has been analyzed.

The major advantage is that full assimilation is only applied on the first temporal window Wi_1, that has a short duration. This assimilation is based on an optimization method [9] and requires, at each iteration, a forward integration of $\mathbb{M}$ and a backward integration of its adjoint [5]. The complexity is proportional to the image size multiplied by the number of time steps in the assimilation window. On the next windows Wi_m, the complexity greatly decreases as the state vector involved in the reduced models is of size $K + L$, which is less than 10 in the experiments.

5. RESULTS

Twin experiments were designed to quantify the sliding windows assimilation method and its ability to process long-term satellite data.

On a first experiment, the full model is run to obtain a four-day simulation, with a time step of 15 minutes and initial conditions displayed on Fig. 1. Snapshots of I_s are taken as observation images $I = \{I^z\}_{z=1...Z}$. Four of them are displayed on Fig. 2. The whole temporal window corresponds

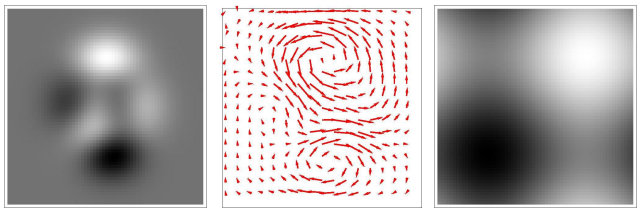
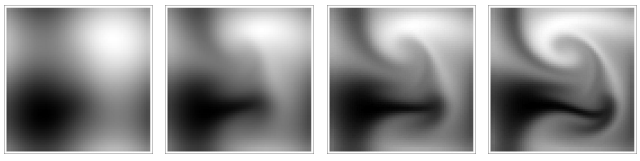


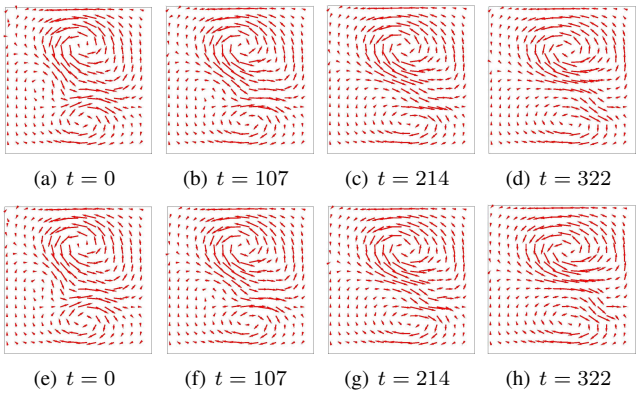
Fig. 1. Left to right: Initial condition $\xi(0)$, $\mathbf{w}(0)$and $I_s(0)$.



(a) $t = 21$ (b) $t = 121$ (c) $t = 221$ (d) $t = 321$

Fig. 2. Observations at temporal indexes 21, 121, 221, 321.

to indices from 0 to 322. It is split in subwindows and assimilation is performed, as described in Section 4, with the sliding windows assimilation method. Results are displayed on Fig. 3. In order to demonstrate the potential of this sliding windows method, we made statistics on the quality of the



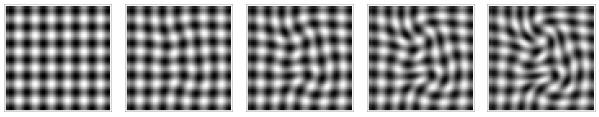
(a) $t = 0$ (b) $t = 107$ (c) $t = 214$ (d) $t = 322$

(e) $t = 0$ (f) $t = 107$ (g) $t = 214$ (h) $t = 322$

Fig. 3. Up: ground truth, Down: estimation with reduced model at temporal indexes 0, 107, 214, 322.

retrieved vorticity on the whole sequence. The NRMSE (in percentage) ranges from 1.1 to 4.0% from the first to the sixth sub-window. The correlation value between the retrieved vorticity and the ground truth decreases from 0.99 to 0.96 from the first to the last sub-window.

A second experiment modifies the initial condition on the pseudo-image, in the simulation providing the snapshots. The full and POD models are quantified on the sub-window with indexes from 0 to 82. Observations are considered at indexes 1, 21, 41, 61, 81, and displayed on Fig. 4. Ground truth

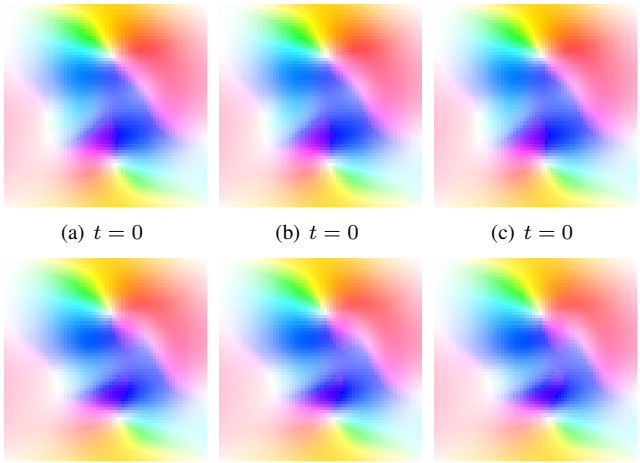


(a) $t = 1$ (b) $t = 21$ (c) $t = 41$ (d) $t = 61$ (e) $t = 81$

Fig. 4. Observations.

and motion results are displayed on Fig. 5. Visualization is made with the colored representation tool of the Middlebury database[1]. On these data, the memory size required by the full model is equal to the size of the image domain multiplied by the size of the temporal window and by the size of the state vector (which is 2), while those of the reduced model is equal to the size of the temporal window multiplied by the size of the reduced state vector (K + L, which is less than 10 in experiments). Results obtained with the reduced model have also been compared with four well-known state-of-the-art methods: [1, 10, 11, 12]. For these methods, optimal parameter values have been used, that provides the smallest errors. In Table 1, the error between the motion result and the ground truth is given for all methods.

[1] http://vision.middlebury.edu/flow/



(a) $t = 0$ (b) $t = 0$ (c) $t = 0$

(d) $t = 81$ (e) $t = 81$ (f) $t = 81$

Fig. 5. Left to right: groundtruth, full model, reduced model.

Table 1. Error analysis: misfit between motion results and ground truth.

Method	Ang. err. (in deg.) Mean	Ang. err. (in deg.) Std. Dev.	Relative norm err. Mean (in %)
Horn[1]	15.26	9.65	45.75
Isambert [10]	10.61	6.92	34.84
Suter[11]	10.41	5.34	37.65
Sun [12]	8.76	4.26	29.07
Full Model	0.18	0.10	0.65
Reduced Model	0.19	0.11	6.50

6. CONCLUSIONS

This paper describes a sliding windows assimilation method, that allows estimating motion on long temporal image sequences, thanks to data assimilation techniques. The method splits the initial temporal window in sub-windows, on which reduced models are computed that allow to process images in quasi-real time. The method is quantified with twin experiments to demonstrate its potential for processing long-term satellite data. The main perspective is to replace the bases Ψ_ξ of the reduced models, which are obtained with a Principal Order Decomposition, by a fixed basis. In that case, even the first sub-window could be processed by a reduced model, in order to further reduce the computational requirements. Moreover, this fixed basis should be defined as satisfying optimality criteria, which translate properties on motion fields and image data. In that way, the method will be able to process long satellite sequences acquired over a full basin, as the Black Sea.

7. REFERENCES

[1] B.K.P. Horn and B.G. Schunk, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[2] I. Cohen and I. Herlin, "Optical flow and phase portrait methods for environmental satellite image sequences," in *Proceedings of European Conference on Computer Vision*, Oxford, UK, April 1996, pp. 141–150.

[3] N. Papadakis and É. Mémin, "Variational optimal control technique for the tracking of deformable objects," in *Proceedings of International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007.

[4] O. Titaud, A. Vidard, I. Souopgui, and F.-X. Le Dimet, "Assimilation of image sequences in numerical models," *Tellus A*, vol. 62, pp. 30–47, 2010.

[5] D. Béréziat and I. Herlin, "Solving ill-posed image processing problems using data assimilation," *Numerical Algorithms*, vol. 56, no. 2, pp. 219–252, Feb. 2011.

[6] K. Drifi and I. Herlin, "Assimilation d'images dans un modèle réduit pour l'estimation du mouvement," in *GRETSI 2011 - Groupe d'Etudes du Traitement du Signal et des Images*, Bordeaux, France, Sept. 2011.

[7] P. Holmes, Lumley J. L., and G. Berkooz, *Turbulence, Coherent Structures, Symmetry and Dynamical Systems*, Cambridge Monographs on Mechanics and Applied Mathematics. Cambridge University Press, 1996.

[8] J. L. Lumley, "The structure of inhomogeneous turbulence," *Atmospheric Turbulence and Radio Wave Propagation*, pp. 166–178, 1967.

[9] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.

[10] T. Isambert, I. Herlin, and J.-P. Berroir, "Fast and stable vector spline method for fluid flow estimation," in *Proceedings of International Conference on Image Processing*, 2007, pp. 505–508.

[11] D. Suter, "Motion estimation and vector splines," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 1994, pp. 939–942.

[12] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *Proceedings of European Conference on Computer Vision*, 2010, pp. 2432–2439.